



Mathematical symbol hypothesis recognition with rejection option

Frank Julca-Aguilar, Nina S. T. Hirata, Christian Viard-Gaudin, Harold Mouchère, Sofiane Medjkoune

► To cite this version:

Frank Julca-Aguilar, Nina S. T. Hirata, Christian Viard-Gaudin, Harold Mouchère, Sofiane Medjkoune. Mathematical symbol hypothesis recognition with rejection option. 14th International Conference on Frontiers in Handwriting Recognition, Sep 2014, Crete, Greece. pp.500 – 504, 10.1109/ICFHR.2014.90 . hal-01096531

HAL Id: hal-01096531

<https://hal.science/hal-01096531>

Submitted on 17 Dec 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Mathematical symbol hypothesis recognition with rejection option

Frank Julca-Aguilar, Nina S. T. Hirata
*Department of Computer Science
 Institute of Mathematics and Statistics,
 University of São Paulo
 São Paulo, Brazil.
 {faguilar;nina}@ime.usp.br*

Christian Viard-Gaudin, Harold Mouchère, Sofiane Medjkoune
*LUNAM LUNAM Université, Université de Nantes
 IRCCyN UMR CNRS 6597 Polytechnique Nantes
 rue Christian Pauc – BP 50609 – 44306, Nantes, France
 {christian.viard-gaudin, harold.mouchere, sofiane.medjkoune}
 @univ-nantes.fr*

Abstract—In the context of handwritten mathematical expressions recognition, a first step consist on grouping strokes (segmentation) to form symbol hypotheses: groups of strokes that might represent a symbol. Then, the symbol recognition step needs to cope with the identification of wrong segmented symbols (false hypotheses). However, previous works on symbol recognition consider only correctly segmented symbols. In this work, we focus on the problem of mathematical symbol recognition where false hypotheses need to be identified. We extract symbol hypotheses from complete handwritten mathematical expressions and train artificial neural networks to perform both symbol classification of true hypotheses and rejection of false hypotheses. We propose a new shape context-based symbol descriptor: *fuzzy shape context*. Evaluation is performed on a publicly available dataset that contains 101 symbol classes. Results show that the fuzzy shape context version outperforms the original shape context. Best recognition and false acceptance rates were obtained using a combination of shape contexts and online features: 86% and 17.5% respectively. As false rejection rate, we obtained 8.6% using only online features.

Keywords—Mathematical symbol classification and rejection; symbol segmentation; shape context.

I. INTRODUCTION

Recognition of handwritten mathematical expressions implies solving three tasks: (1) symbol segmentation, (2) symbol recognition and (3) structural analysis. In the first task, strokes that belong to a same symbol must be grouped. In the second task, a label must be assigned to each symbol; and in the structural analysis task, relations between symbols must be identified, for example superscript or subscript relations.

As it is difficult to find a correct segmentation, several symbol hypotheses are generated and classifiers must give a recognition confidence to each hypothesis. Most symbol classifiers are trained on the recognition of isolated symbols (only true hypothesis) [1], [2], [3], and it is assumed that confidence values of miss-segmented symbols (false hypotheses) will be lower than those for the correctly segmented symbols.

A different approach uses both true and false hypotheses to explicitly train the classifier to identify or reject mis-segmented symbols [4]. Previous results suggest that training classifiers with this approach can improve the performance

of a mathematical expression recognition system [4]. However, to the best of our knowledge, there are no previous works that focus on the development of these classifiers with rejection option.

In this work, we focus on the problem of handwritten mathematical symbol hypothesis recognition with rejection option. We extract symbol hypotheses from complete handwritten mathematical expressions and train artificial neural networks to perform symbol classification of true hypotheses and rejection of false hypotheses (Section II). In addition, we propose a new version of the well-known shape context descriptor (SC) [5], called: *fuzzy shape context* (F-SC) (Section III). To provide comparable results (Section IV), we evaluate the proposed methods on the publicly available Competition on Recognition of On-line Mathematical Expressions (CROHME 2013) dataset [6].

II. MATHEMATICAL SYMBOL HYPOTHESIS RECOGNITION AND REJECTION

An online handwritten mathematical expression can be represented as a sequence of strokes:

$$E = (s_1, s_2, s_3, \dots, s_t) \quad (1)$$

where s_i is the i -th written stroke, considering time order.

For an expression composed of t strokes, the total number of symbol hypotheses that can be generated is $O(2^t)$. To limit the number of symbol hypotheses, several heuristics have been applied: limiting the number of strokes (between 4 or 5) that form a symbol [4], [1], considering only groups of consecutive (in time order) strokes [7], [8], and considering groups of only intersected strokes [9].

Despite the fact that the proposed heuristics reduce the number of symbol hypotheses, the number of false hypotheses is generally bigger than the true ones.

Handwritten variability makes difficult to differentiate true hypotheses from the false ones. According to the CROHME 2013 competition [6], the best recall and precision of segmentation obtained by the best University system were 84.97% and 87.08%, respectively.

Recognition of a large number of mathematical symbol classes is by itself another difficult problem. In addition to

the intrinsic similarity between some symbols (such as C, c, P, p) the handwritten nature introduces more ambiguity between symbols. Results in the CROHME 2013 competition suggest that there are still a considerable room for further improvements in terms of symbol classification.

As we do not consider that symbols are already segmented, the scope of our problem includes, at some extent, both symbol segmentation and recognition. The symbol segmentation is given by the rejection of false hypotheses, and the symbol classification is given by the classification of the correctly accepted symbol hypotheses.

To generate symbol hypotheses from mathematical expressions, we use a symbol hypothesis generator tool provided in the Fourth International Competition on Handwritten Mathematical Expression Recognition (CROHME-IV)¹. Given an expression, this tool allows us to extract the whole set of true symbol hypotheses and also false symbol hypotheses. The false hypotheses are generated by grouping consecutive (in time order) strokes that not represent true symbols in the expression.

III. FEATURE SET

We extract offline and online features from symbol hypotheses. As offline features, we use the previously defined shape context [5] and propose a new shape context version: *fuzzy shape context*. As online features, we use a set of raw online data.

A. Shape Context

Shape context was proposed as a shape descriptor in [5] and has been applied in several symbol recognition problems with outstanding results. Applications include recognition of handwritten digits and 3D objects [5], handwritten Tamil scripts with 156 symbol classes [10] and leaf image classification (220 leaf classes) [11].

Given a set of points $P = \{p_1, p_2, \dots, p_n\}$, the shape context of a point p_i in P is a log polar histogram that expresses the distribution of the remaining points relative to p_i [5]. Figure 1 illustrates shape context calculation for two points of a symbol. In that example, the symbol area is divided into 8 angular regions and 3 radial regions, for a total of 24 bins. Each bin contains the quantity of points spatially placed in that bin.

Given that a shape context histogram is defined over a log polar space, the shape context relative to a point p_i is more sensitive to positions of nearby sample points (local features) than those of points farther away (global features) [5].

B. Fuzzy shape context

The shape context descriptor was originally defined over crisp bins. Considering this definition, each point belongs only to one bin of an histogram.

¹<http://www.isical.ac.in/~crohme/>

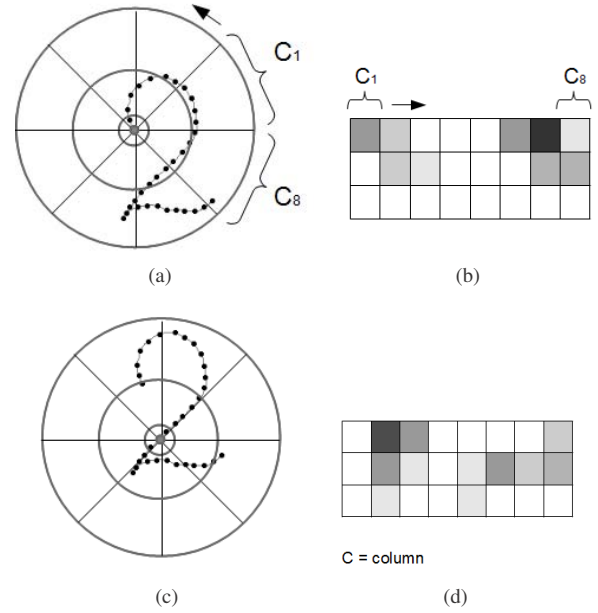


Figure 1. Shape context of two points of a symbol “2”: (a) and (c) show the sampled points and log polar histogram bins used to calculate shape context. (b) and (d) show the shape context histogram relative to (a) and (c) respectively. Dark cells mean higher values.

During the sampling process or due to handwritten variability, small changes in sampled points may be generated. These changes may affect specially points falling near to the limits of bins: small displacements near to the end of a bin may change the total of number of points in that bin.

We extend the shape context definition by considering bins as fuzzy sets, as shown in figure 2.

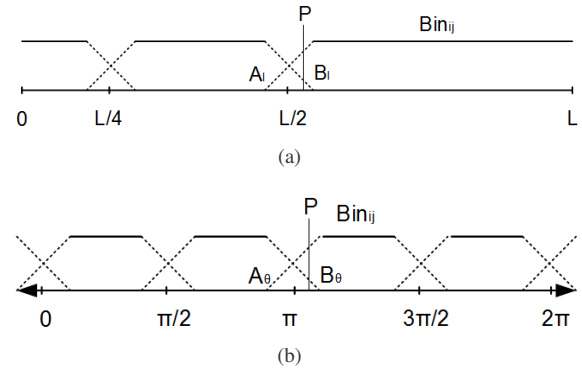


Figure 2. Fuzzy bins in the (a) radial and (b) angular coordinates. The arrows in the angular coordinates indicate the continuation given by the circular nature of shape context.

Given a shape context histogram $H = \{Bin_{ij}\}$, where i indexes the bins relative to radial coordinate and j indexes bins in the angular coordinate. For a point $P = (P_l, P_\theta)$, that lies in a bin Bin_{ij} , the membership value of point P

is given by:

$$M_{ij}(P) = \alpha_i * \beta_j \quad (2)$$

where α_i and β_j are between 0 and 1 and indicate the confidence value of P lying in Bin_{ij} relative to the radial and angular coordinates, respectively. When P lies in an intercepting region, the values α_i and β_j are calculated according to its position relative to the corresponding coordinate, otherwise they take the value 1. For example, for the point in P in figure 2, the confidence values will be:

$$\alpha_i = \frac{P_l - A_l}{B_l - A_l} \quad (3)$$

$$\beta_j = \frac{P_\theta - A_\theta}{B_\theta - A_\theta} \quad (4)$$

Note that with these new definitions, depending on its position, a point may belong up to four bins. Figure 3 shows four different positions for calculating the membership values:

- P position, only bin Bin_{ij} receives a non-zero value (=1),
- R position, bin Bin_{ij} and a radial-connected bin receive non-zero values,
- S position, bin Bin_{ij} and an angular-connected bin receive non-zero values,
- Q position, bin Bin_{ij} and three connected bins receive non-zero values

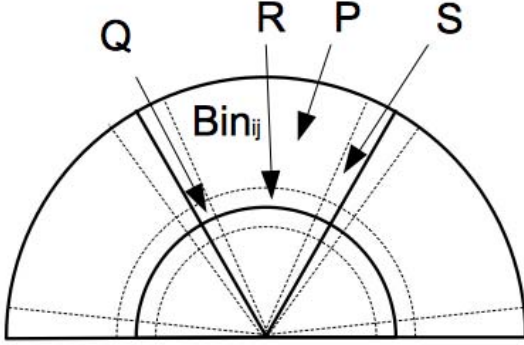


Figure 3. Four different positions in a bin Bin_{ij} . Regions between dotted lines indicate transition areas between bins

C. Shape contexts as input features for Neural Networks

Shape contexts has generally been applied using a shape matching approach: given two sets of points $P = \{p_1, \dots, p_n\}$ and $Q = \{q_1, \dots, q_n\}$, to calculate a similarity between P and Q , shape context is calculated at each point and a best matching between points of P and Q is calculated, using the χ^2 metric [5].

However, in a previous work, we showed that using shape contexts as input features for neural networks, results are

comparable to the matching based approach, but with a considerable efficiency improvement [12].

When using shape contexts within neural networks, we can reduce number of features without decreasing performance. As shape contexts of near points tend to be similar, we do not need to calculate shape contexts at each sampled point. In contrast, we can keep the rate of sampled points (for instance, 30 points per symbol), but extract shape contexts of only a subset of points. For example, when using online data, shape contexts of some consecutive (in time) points could be avoided.

In our shape context implementation, we used 3 radial regions and 8 angular regions, for a total of 24 bins. We extracted 30 points per symbol but only calculated shape contexts of eight equally distributed (in time order) points. These values were determined in a previous work [12]. Accordingly, the size of a SC and F-SC feature vector is $8 \times 24 = 192$.

D. Online Features

As mentioned above, we also extracted raw online data. For each sampled point, 7 features are extracted: (2) normalized coordinates, sin and cousin of the direction and curvature (4) and a binary value (0 or 1) that indicates the state of the stylus in the point [13]. As in this case we also extract 30 points per symbol, the dimension of the feature vector is $30 \times 7 = 210$.

IV. EXPERIMENTATION

A. Experimental setup

To evaluate the proposed methods, we used the CROHME 2013 dataset² [6]. This dataset includes expressions collected from several laboratories around different countries through several kinds of input devices, as digital pen technologies, white-boards and tablets with sensible screens. Given this variety of devices, symbols were sampled in different scales and resolutions.

The dataset is divided into a training part with 8,835 mathematical expressions and 85,803 symbols and a test part with 671 expressions and 5,889 symbols. The number symbol classes is 101.

Using the CROHME-IV symbol generator tool, we generated 74,285 false symbol hypothesis from the training expressions and 5,276 from the test expressions. Thus, in our evaluation, the ratio between true and false hypothesis is near 1.

It is important to note that symbols which are considered here come from entire mathematical expressions. Hence, a symbol might be influenced by its context. For example, symbols written before and after a symbol X may influence the way people write X. This may have introduced more variability on symbol classes.

²The CROHME 2013 dataset is publicly available at: http://www.iapr-tc11.org/mediawiki/index.php/Datasets_List.

For training neural networks, we split the training data and perform a cross-validation scheme. In this scheme, we selected a Multi Layer Perceptron neural network with one hidden layer containing 150 units when the two sets of features are used separately. When both feature sets are used jointly, we also selected one hidden layer but, in this case, it is composed of 200 units.

We evaluate our methods using symbol recognition rate (SRR), false acceptance rate (FAR) and false rejection rate (FRR). Considering Table I, FAR and FRR are calculated as:

$$FAR = \frac{\#\{\text{Accepted junk}\}}{\#\{\{\text{Accepted junk}\} \cup \{\text{Rejected junk}\}\}} \quad (5)$$

$$FRR = \frac{\#\{\text{Rejected symbol}\}}{\#\{\{\text{Accepted symbol}\} \cup \{\text{Rejected symbol}\}\}} \quad (6)$$

Table I
SYMBOL HYPOTHESES CLASSIFIER OUTCOMES FOR FAR AND FRR

Output	Actual class	
	Symbol	Junk
Accepted	Accepted symbol	Accepted junk
Rejected	Rejected symbol	Rejected junk

From the $\{\text{Accepted symbol}\}$ set, we determine the symbols that have been correctly classified (between the 101 classes): $\{\text{Correct classification}\}$. Thus, we calculate SRR as:

$$SRR = \frac{\#\{\text{Correct classification}\}}{\#\{\text{Accepted symbol}\}} \quad (7)$$

In the context of a mathematical recognition system, low FAR values allow us to reduce the search space for the correct segmentation and classification search. On the other hand, low FRR values allow us to keep more true hypotheses for a posterior final recognition, for example using the complete expression structure.

B. Results and discussion

Results relative to the Top-1, Top-2 and Top-3 recognition rates for the test set are presented in table II. Regarding to the features used separately, we can see that the F-SC features are the best, with 84.38% Top-1 recognition rate. Considering all features set, the best results were obtained combining SC and F-SC with online features: around 86% Top-1 recognition rate in both cases.

It is important to note that some symbols can not be recognized by considering only the symbol information. For example, symbols x , X and \times (“\times” latex code) may be handwritten with an identical shape. To solve those ambiguities, context from the mathematical expressions must be used. In these cases, a set of best classifications or top classifications can be used: best possible classifications can

be kept and in the structural analysis step the best class can be selected, considering relations with other symbols. Top-3 recognition rates show that combining both features we can get high confidence of including the true symbol between the possible classes.

Table II
TOP-1, TOP-2 AND TOP-3 RECOGNITION RATES.

Feature set	SRR (%)		
	Top-1	Top-2	Top-3
SC	83.59	90.00	94.07
F-SC	84.38	89.87	94.45
Online features	82.55	90.15	94.20
Online + SC	86.02	92.47	96.07
Online + F-SC	85.99	92.29	96.13

Table III shows results relative to FAR and FRR. In this case, best FAR was obtained by the Online + SC features (17.5%) and best FRR by the online features (8.60%). We can see also that the F-SC version outperforms SC: while the F-SC gets almost the same FAR as the SC, the first has 3% less FRR than the second one.

Note that the FAR and FRR are independent of the ratio symbol/junk in the test set. Although the ratio in the train set (about 1) influence these rates via the training of the classifier, we can see that the obtained FRR and FAR values are far from 50%. This means that our classifier performs much better than a simple random decision with prior probability knowledge.

Table III
FALSE ACCEPTANCE RATE (FAR) AND FALSE REJECTION RATES (FRR).

Feature set	FAR (%)	FRR (%)
SC	20.66	13.47
F-SC	20.72	10.94
Online features	19.45	8.60
Online + SC	17.5	11.26
Online + F-SC	18.39	11.23

V. CONCLUSIONS AND FURTHER WORK

In this work, we propose a mathematical symbol hypothesis classifier with rejection option. As part of the features set, we propose a new shape context-based feature: *Fuzzy Shape Context*. In contrast to the common matching based approach, we used shape contexts as input features for neural networks. In this architecture, the F-SC outperforms the SC in terms of recognition rate and FRR. Best recognition rates and FAR were obtained combining both features and best FRR was obtained using only online features.

The good performance of shape context and its combination with online data suggests further research in this area. At this regard, an important issue that can be studied is the way we extract shape context. As described above, in the current approach, we extract shape contexts at eight

equally distributed (in time order) points. As a result, shape contexts is also influenced by the online information: the order in which strokes are introduced or writing direction may change the order in which shape contexts are input to the neural networks. To obtain *totally* offline features, we could extract shape contexts at some key points, for example, the center bounding box of a symbol.

The proposed classifier and future improvements will be integrated and evaluated into a handwritten mathematical expression recognition system.

REFERENCES

- [1] N. E. Matsakis, "Recognition of handwritten mathematical expressions," Master's thesis, Massachusetts Institute of Technology, Cambridge, 1999.
- [2] S. MacLean and G. Labahn, "A new approach for recognizing handwritten mathematics using relational grammars and fuzzy sets," *International Journal on Document Analysis and Recognition (IJDAR)*, vol. 16, no. 2, pp. 139–163, 2013.
- [3] F. Álvaro, J.-A. Sánchez, and J.-M. Benedí, "Recognition of on-line handwritten mathematical expressions using 2d stochastic context-free grammars and hidden markov models," *Pattern Recognition Letters*, no. 0, pp. 58 – 67, 2012.
- [4] A.-M. Awal, H. Mouchère, and C. Viard-Gaudin, "A global learning approach for an online handwritten mathematical expression recognition system," *Pattern Recognition Letters*, vol. 35, no. 0, pp. 68 – 77, 2012.
- [5] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, pp. 509–522, April 2002.
- [6] H. Mouchere, C. Viard-Gaudin, U. Garain, D. H. Kim, K. J. H., and Z. R. (2013, April) Icdar 2013 crohme: Competition on recognition of online handwritten mathematical expressions.
- [7] B. Huang and M.-T. Kechadi, "A structural analysis approach for online handwritten mathematical expressions," in *International Journal of Computer Science and Network Security*, 2007.
- [8] S. Lehmberg, H.-J. Winkler, and M. Lang, "A soft-decision approach for symbol segmentation within handwritten mathematical expressions," in *International Conference on Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings, 1996 IEEE*, vol. 6, May 1996, pp. 3434–3437 vol. 6.
- [9] E. Tapia, "Understanding mathematics: A system for the recognition of on-line handwritten mathematical expression," Ph.D. dissertation, Freie Universitt Berlin, 2004.
- [10] L. Prasanth, V. Babu, R. Sharma, G. V. Rao, and D. M., "Elastic Matching of Online Handwritten Tamil and Telugu Scripts Using Local Features," in *Proceedings of the Ninth International Conference on Document Analysis and Recognition - Volume 02*, 2007, pp. 1028–1032.
- [11] Z. Wang, B. Lu, Z. Chi, and D. Feng, "Leaf image classification with shape context and sift descriptors," in *Digital Image Computing Techniques and Applications (DICTA), 2011 International Conference on*, 2011, pp. 650–654.
- [12] F. Julca-Aguilar, C. Viard-Gaudin, H. Mouchère, S. Medjkoune, and N. Hirata, "Integration of shape context et neural networks for symbol recognition," in *Semaine du Document Numérique et de la Recherche d'Information 2014 (SDNRI)*, 2014.
- [13] A.-M. Awal, "Reconnaissance de structures bidimensionnelles : Application aux expressions mathématiques manuscrites en ligne," Ph.D. dissertation, Image and VideoCommunication, research group, IRCCyN, 2010.